Full-waveform inversion using alternative objective functions in the presence of noise and uncertainties of source signature

Leonardo Gómez, CPGG/UFBA & Reynam C. Pestana*, CPGG/UFBA

Abstract

Full-waveform inversion (FWI) is a method able to estimate model parameters in subsurface from seismic data. The algorithm consists on the minimization of an objective function that relates observed seismic data and synthetic data for the estimated model. Thus, the successful application of FWI depends on the accurate correspondence between modeled and field data.

Some problems that appear in the application of FWI to field data are dealt by using alternative functionals for objective function. The results, obtained by using synthetic data, show that the definition based on the least absolute value norm and the cross-correlation error measure have advantages over least squares norm, turning the method more robust when facing noisy data and uncertainties in source signature used for inversion.

The quasi-Newton formulation of L-BFGS method is applied to non-quadratic objective functions without affecting, in practice, the convergence of inversion scheme.

Introduction

Full-waveform inversion (FWI) is conceived as a procedure in which synthetic seismic data, modeled using an estimated model, is matched to field data, fitting all the observable properties of seismic data: recording time, amplitude and phase (Tarantola, 1984; Lailly, 1983). The scheme is guided by an objective function that is minimized at each iteration and measures, in some sense, the mismatch between those data vectors.

One of the main challenges of the application of FWI in field data is the definition of minimization criteria presenting a stable performance in the presence of amplitude errors in seismic data (Virieux and Operto, 2009). Least-squares are commonly used for solving inverse problems. Although computations based on this formulation are simple, solution of the problem lacks of robustness, being sensitive to aleatory outliers in the data set due to the assumption that all initial uncertainties in the problem can be modeled using Gaussian distributions (Tarantola, 2005).

Considering alternative formulations for objective functions defined in terms of data amplitude, the influence of non coherent noise can be diminished, such as p-norm for error between data vectors, for $p \neq 2$, and specifically l_1-norm. This norm is not based on the hypothesis of Gaussian uncertainties making its results sufficiently insensitive to outliers affecting data (Tarantola, 2005).

On the other hand, seismic perturbation that acts as source is generally not well determined and has to be considered as an unknown in the inversion process. This way, source term can be estimated and updated in FWI algorithm, alternating with model parameters (Tarantola, 1984). Otherwise, can be defined objective functions that turn inversion process independent from source function or less sensitive to errors in its approximation.

In this work, we review three formulations for FWI objective function: l_2-norm, l_1-norm and cross-correlation of data vectors. We describe their properties and their variation with errors in data. Through synthetic experiments, in which we have induced some problems that affect the application of the algorithm in real data, we show how alternative approaches for objective function can make the scheme less sensitive to noisy data and uncertainties in estimated source signature.

Theory

Modeling problem

Considering an isotropic and homogeneous physical medium, with constant density, the propagation of an acoustic oscillatory perturbation is completely describe, in both space an time, by the two-way wave equation:

$$\frac{1}{c(\mathbf{x})^2}\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \nabla^2 u(\mathbf{x},t) = s(\mathbf{x},t)\delta(\mathbf{x}-\mathbf{x}_s), \qquad (1)$$

where $u(\mathbf{x},t)$ represents the seismic wavefield at time t and for position $\mathbf{x} = (x,y,z)$. $c(\mathbf{x})$ is the velocity of propagation of the acoustic wave. Source term, acting on the position $\mathbf{x}_s$, is denoted by $s(\mathbf{x},t)$ and ∇^2 corresponds to Laplacian operator in Cartesian coordinates. Initial conditions for this differential problem are usually stated as $u(\mathbf{x},t=0)=0$ and $\partial u(\mathbf{x},t=0)/\partial t = 0$. In this work, we use the rapid expansion method (REM) as extrapolation operator (Pestana and Stoffa, 2010).

Objective function

The non-linear inversion problem is formulated as the problem of estimating the wavefield $\mathbf{u}$ and the model vector $\mathbf{m}$ that satisfy the forward modeling equation (1), such that the model is the closest to initial model and the distance between synthetic data vector $\mathbf{d}_{\text{cal}}(\mathbf{m})$ and observed data vector $\mathbf{d}_{\text{obs}}(\mathbf{m})$ is minimum (Tarantola, 1984). The definition of distance is established through a functional called objective function.

Vector $\mathbf{m}$ represents the physical parameters of the discrete subsurface that describe propagation phenomena and, in our case, correspond to the values of compressional wave velocity at position $\mathbf{x} = (x, y, z)$. Observed data vector $\mathbf{d}_{\text{obs}} = d_{\text{obs}}(\mathbf{x}_r, t; \mathbf{x}_s)$ contains a set of wavefield measurements at discrete receiver positions $\mathbf{x}_r$ associated to a single seismic source located at $\mathbf{x}_s$, registered at discrete time intervals t. Elements of calculated data vector $\mathbf{d}_{\text{cal}} = d_{\text{cal}}(\mathbf{x}_r, t; \mathbf{x}_s)$ are obtained from the modeled wavefield $u(\mathbf{x}, t)$ on current model by using an operator $\mathscr{R}$ that extracts wavefield values for all discrete time steps, at receiver positions $d_{\text{cal}}(\mathbf{x}_r, t; \mathbf{x}_s) = \mathscr{R}u(\mathbf{x}, t)$, for each source (Virieux and Operto, 2009).

Objective function, minimized in FWI iterative process, is a relation $E(\mathbf{m}) : \mathbf{M} \rightarrow \Re$ that compares registered and modeled data vectors, for each source-receiver pair in seismic array and measures, in some sense, the mismatch between them. Besides l_2-norm, we review l_1-norm and correlation between data vectors to define distance measure in data space.

Least-squares norm

The conventional implementation of waveform inversion is based on least-squares or l_2 norm, which is defined as the difference in amplitude between predicted and observed data as objective function (Tarantola, 1984)

$$E_{l_2}(\mathbf{m}) = \frac{1}{2}\|\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}}\|_2^2 = \frac{1}{2}\sum_{s,r}\int_0^T (d_{\text{cal}} - d_{\text{obs}})^2 dt, \quad (2)$$

being T the record length. Summing is performed over source-receiver pairs s-r in seismic array and total number of time samples in seismograms.

The approach described by least squares norm emphasizes mainly on correlating record amplitude, assuming that the error distribution is Gaussian-like. In the cases in which this assumption is not satisfied, e.g. when atypical isolated values affect data amplitude, inversion process turns less robust (Virieux and Operto, 2009).

Least-absolute norm

Alternatively, absolute value of the difference in amplitude between synthetic and observed data, or l_1-norm, can be adopted for the definition of $E(\mathbf{m})$ (Brossier et al., 2010)

$$E_{l_1}(\mathbf{m}) = \|\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}}\|_1 = \sum_{s,r}\int_0^T |d_{\text{cal}} - d_{\text{obs}}| dt. \quad (3)$$

This norm is not based on Gaussian statistics in data space and is considered little sensitive to noise.

Least-absolute norm is not differentiable at $\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}} = \mathbf{0}$, causing singularity in gradient vector at the minimum corresponding to an optimal model that explains observed data completely. Even that this situation is unlikely in real applications, using l_1-norm can derive in some complications in the near vicinity of global minimum (Brossier et al., 2010). Some alternative strategies propose starting the inversion process with an objective function defined in terms of l_1-norm and pass to l_2-norm as model gets close to optimal solution. This approach is treated by some formulations such that Huber criterion (Huber, 1973) and hybrid l_1/l_2 criterion (Bube and Langan, 1997).

Cross-correlation error

A normalized cross-correlation based objective function can be formulated as a measure of similarity between modeled and observed data vectors (Zhang et al., 2015; Klimm, 2013)

$$\begin{aligned} E_c(\mathbf{m}) &= -\sum_s \frac{\langle \mathbf{d}_{\text{cal}}, \mathbf{d}_{\text{obs}} \rangle}{\|\mathbf{d}_{\text{cal}}\|_2 \|\mathbf{d}_{\text{obs}}\|_2} \\ &= -\sum_s \frac{\sum_r \int_0^T d_{\text{cal}} d_{\text{obs}} dt}{\left(\sum_r \int_0^T d_{\text{cal}}^2 dt\right)^{1/2} \left(\sum_r \int_0^T d_{\text{obs}}^2 dt\right)^{1/2}}, \end{aligned} \quad (4)$$

where $\langle \cdot, \cdot \rangle$ represents scalar inner product. Considering the dataset recorded at a single seismic experiment, when synthetic and registered are co-linear and have the same orientation in data vector space, the objective function (4) presents the minimum value, equal to -1. For any other situation, its value ranges from -1 to 1.

The objective function based on cross-correlation is sensitive to similarity between synthetic and observed data, only; it emphasizes mainly on matching the phase in seismic data, being more flexible to amplitude correspondence required by l_2-norm definition. FWI formulated using this objective functional is equivalent to phase inversion in time domain, in which mismatch in phase between modeled and recorded data is minimized (Dutta et al., 2014). This approach is useful when synthetic data amplitude do not match with that of observed data.

Inversion scheme

Conventionally, a local gradient-based iterative scheme is used for FWI. Current model vector $\mathbf{m}_k$ is updated via (Nocedal and Wright, 2006)

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{p}_k, \quad (5)$$

where, $\mathbf{p}_k$ stands for the search vector in the direction of the minimum of the objective function in the neighborhood of $\mathbf{m}_k$. The step-length α_k scales the search direction before updating the model. Search direction can be estimated using the approach stated by steepest-descent method, such that updated model is searched in the opposite direction given by the gradient vector, i.e., $\mathbf{p}_k = -\nabla E(\mathbf{m}_k)$, or through quasi-Newton formulations, in which $\mathbf{p}_k \approx -\mathbf{H}_k^{-1}\nabla E(\mathbf{m}_k)$, such as L-BFGS algorithm (Nocedal, 1980; Nocedal and Wright, 2006).

Methods of Newton and quasi-Newton classes can be applied to quadratic or locally quadratic objective functions, since they require this function is continuous and twice differentiable for ensuring convergence of the algorithm. Brossier et al. (2010) showed that, in practice and even without satisfying this conditions, L-BFGS method can be applied to inverse problems using objective functions based on l_1-norm, without affecting convergence significantly. The results of this work indicate that this conclusion is also valid for the formulation based on cross-correlation between data vectors.

Gradient vector, required for local-scope inversion schemes, is conventionally computed as (Tarantola, 1984)

$$\nabla E(\mathbf{m}) = \frac{2}{c(\mathbf{x})^3}\sum_s \int_0^T \lambda(\mathbf{x}, t)\frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} dt, \quad (6)$$

assuming a point-collocation scheme for model parametrization. It contains the local sensibility information of objective function with respect to each model parameter. In the context of adjoint-state method (Plessix, 2006) $u(\mathbf{x},t)$ corresponds to the state variable, and $\lambda(\mathbf{x},t)$ to the adjoint-state variable, obtained by solving the reverse-time modeling problem

$$\frac{1}{c(\mathbf{x})^2}\frac{\partial^2\lambda(\mathbf{x},t)}{\partial t^2} - \nabla^2\lambda(\mathbf{x},t) = \frac{\partial E}{\partial \mathbf{u}}(\mathbf{x},t), \qquad (7)$$

with final conditions $\lambda(\mathbf{x},t=T)=0$ and $\partial\lambda(\mathbf{x},t=T)/\partial t = 0$. The source term in (7) is called the virtual secondary source. It is a composite source, with an elementary source at each receiver position, and represents a form of residuals between synthetic and observed data. Depending on the formulation of objective function, it is calculated, for a single seismic source, as (Tarantola, 1984; Brossier et al., 2010; Zhang et al., 2015)

$$\frac{\partial E_{l_2}}{\partial \mathbf{u}}(\mathbf{x},t) = \sum_r (\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}})\,\delta(\mathbf{x}-\mathbf{x}_r), \qquad (8a)$$

$$\frac{\partial E_{l_1}}{\partial \mathbf{u}}(\mathbf{x},t) = \sum_r \frac{\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}}}{|\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}}|}\,\delta(\mathbf{x}-\mathbf{x}_r) \qquad (8b)$$

and

$$\frac{\partial E_c}{\partial \mathbf{u}}(\mathbf{x},t) = \sum_r \frac{1}{\|\mathbf{d}_{\text{cal}}\|_2\|\mathbf{d}_{\text{obs}}\|_2} \times \left[\frac{\langle \mathbf{d}_{\text{cal}}, \mathbf{d}_{\text{obs}}\rangle}{\|\mathbf{d}_{\text{cal}}\|_2^2}\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}}\right]\delta(\mathbf{x}-\mathbf{x}_r), \qquad (8c)$$

for the functionals considered in this work. Gradient has the same form for different objective functions with distinct secondary source. Thus, objective function and gradient vector can be computed at the same computational cost, regardless of the formulation used for $E(\mathbf{m})$ (Brossier et al., 2010).

Step length α_k is computed by quadratic interpolation for steepest-descent method and assumed as unitary for L-BFGS approach. Finally, step length is validated using first Wolfe condition in a backtracking line-search iterative method (Nocedal and Wright, 2006), for ensuring decrease in the value of the objective function.

Numerical Results

In order to illustrate the behavior of the formulations for objective function considered in this work, we designed an experiment in which observed data vector is formed by two elements, $\mathbf{d}_{\text{obs}} = (1,1)$. Parameter values range from -4 and 6, such that maximum mismatch in data Δd_i is 5. Figure 1 shows the behavior of objective function with respect to error between calculated and observed data $\Delta\mathbf{d}$. Comparison of Figures 1.a and 1.c evidences the distinction in sensibility of E_{l_2} and E_{l_1} objective functions. Being less sensitive to large errors in data space, l_1-norm formulation in potentially more robust when compared with l_2-norm. The singularity at $\Delta\mathbf{d} = \mathbf{0}$ in E_{l_1} can also be observed in Figure 1.c. On the other hand, the objective function defined in terms of correlation of data vectors (Figure 1.e) presents its minimum value equal to -1 at all points in which $\mathbf{d}_{\text{cal}} = n\mathbf{d}_{\text{obs}}$, being n a positive scalar. That set of vectors will be solution of the problem of minimizing E_c, given that they satisfy the similarity or parallelism condition imposed by that functional, without considered the mismatch in amplitude. Contrary, E_c takes the maximum value for the set of observations for which $\mathbf{d}_{\text{cal}} = -n\mathbf{d}_{\text{obs}}$.

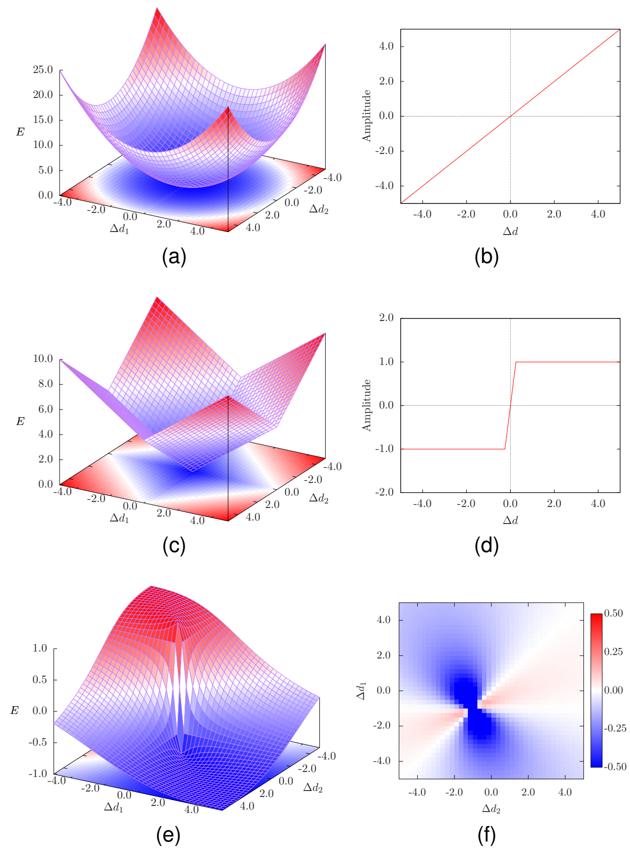

Figure 1: Form of objective function and amplitude of secondary virtual source. The functionals are defined in terms of the l_2-norm (a-b), l_1-norm (b-c) and cross-correlation of data vectors (e-f).

A representation of the amplitude of secondary virtual sources as a function of data residuals for different objective function formulations, for the problem described previously, are shown in Figure 1. For the objective function based on l_2-norm (Figure 1.b), amplitude varies with magnitude of data mismatch, reducing robustness of inversion when big and incoherent outliers affect the data. On the other hand, l_1-norm (Figure 1.d) completely ignores the magnitude of difference in data amplitude when constructing secondary virtual source, making this criterion less sensitive to considerable errors in data. In this case, secondary virtual source corresponds to sign function of the difference in amplitude between the elements of $\mathbf{d}_{\text{cal}}$ and $\mathbf{d}_{\text{obs}}$ (Brossier et al., 2010). When objective function is based on cross-correlation (Figure 1.f), secondary virtual source is scaled accordingly to similarity between observed and synthetic data vectors. Along main diagonal, in which $\mathbf{d}_{\text{cal}} = n\mathbf{d}_{\text{obs}}$, for any scalar n, gradient vector is horizontal; therefore, amplitude of secondary virtual source is null. Sign of amplitude and direction of data error are related. Amplitude is predominantly positive in those cases in which residuals have similar directions and negative when they are opposite. This way, model corrections estimated by gradient vector will tend to make data vectors more similar, or parallel in data space, with iterations. Values smaller than -0.5 were clipped for visualization purposes.

Alternative objective functions are applied for estimating compressional velocity using the synthetic Marmousi model (Figure 2.a). It consists on a grid of 375×369 points, with spacing of 8 m in depth and 25 m in horizontal direction. Observed data were generated using rapid expansion method, assuming a zero-phase Ricker wavelet with dominant frequency of 15 Hz as seismic source. All grid nodes at surface are considered as receivers, for 62 source points equally spaced 150 m. Time sampling interval is 4 ms, comprising 3.0 s of record length. Absorbing boundary condition at surface is considered.

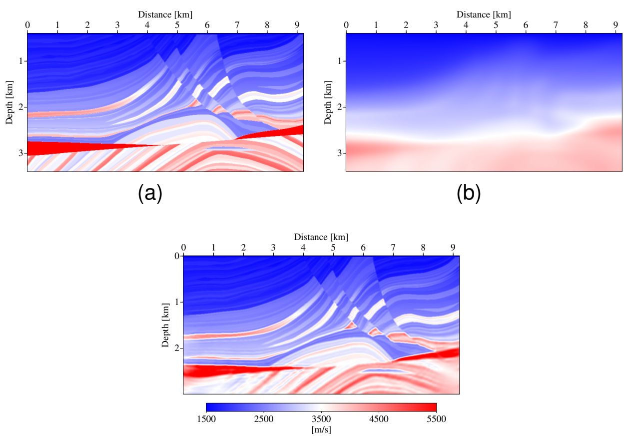

Figure 2: Marmousi model (a), initial model (b) and estimated model using a conventional multi-frequency inversion scheme, defining objective function in terms of l_2-norm (c).

Initial model used as input in iterative scheme (Figure 2.b) corresponds to a smoothed version of true model. In order to avoid the influence of direct arrival, a 24 m water layer is include in the upper part of initial model; velocity in this layer is assumed to be known and is not updated with iterations. Inversion is performed using a multi-frequency hierarchical approach (Bunks et al., 1995). Four inversion stages limited by maximum frequencies equal to 5, 10 , 20 and 30 Hz are considered in the process and 30 iterations are performed at each stage, totalizing 120 iterations. Steepest-descent approach is used at first iteration of each stage. In other case, update direction is computed via L-BFGS method, considering a maximum of 10 model-gradient pairs in the algorithm.

In the firs experiment, synthetic noise-free data are used as input (Figure 3.a) and source signature is assumed to be known. Estimated model, considering a formulation based on l_2-norm for objective function, is shown in Figure 2.c. Recovering of model features is satisfactory. Similar results are obtained for other functionals. The following tests tend to evaluate the behavior of the described objective functions in the FWI scheme when facing factors that affect the application in field data.

Noisy data

In this case, synthetic observed data were contaminated with random noise in a proportion of 20% (Figure 3.b). Results of inversion for objective functions described previously are shown in Figure 4. Even at the presence of noise in observed data, inversion scheme is able to recover the most noticeable features in the model. Quality is reduced at deeper parts where geometry of reflectors is distorted, being evident in salt intrusions and the reservoir at antiform structure. Interval velocity inside stratigraphic layers is also affected by noise.

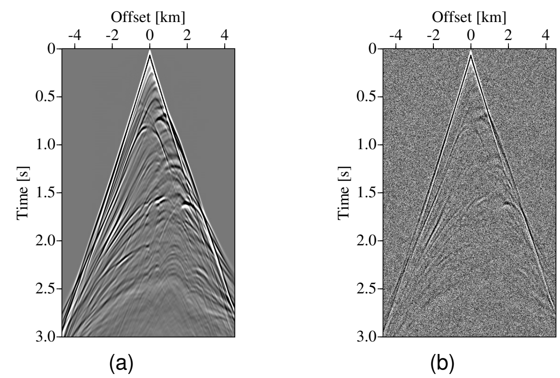

Figure 3: Synthetic data used for inversion before (a) and after (b) adding noise. Seismograms correspond to source locate at $\mathbf{x}_s = (4675, 0)$

An useful measure for evaluating the correctness of estimated model is the normalized root mean square error (NRMS), given by

$$\text{NRMS} = \frac{1}{\max(\mathbf{m}) - \min(\mathbf{m})} \sqrt{\frac{1}{M} \sum_{i=1}^{M} (\mathbf{m}_i - \mathbf{m}_i^{\text{est}})^2}, \qquad (9)$$

where, M denotes the number of parameters in model vector, $\mathbf{m}^{\text{est}}$ corresponds to inverted model, $\mathbf{m}$ is the true model and $\min(\mathbf{m})$ and $\max(\mathbf{m})$ are the minimum and maximum values of model parameters, respectively.

NRMS error in estimated model when using free-noise data is 3.85%, 5.03% and 3.81% for the l_2-norm, l_1-norm and correlation mismatch formulations used for objective function, respectively. When noisy data are used as input, the corresponding NRMS errors are 6.11%, 6.46% and 6.06%. It represents an increment of 58.7%, 28.4% and 59.1% in the value of NRMS error for each functional showing that, using an objective function based on the l_1-norm, the relative behavior of inversion process faced with noise affecting observed data is more robust, when compared with the case of free-noise data.

Robustness of l_1-norm formulation is also observable in the profile of estimated model shown in Figure 5. Although all the three formulations achieve an acceptable result by recovering the main tendency of the values of model parameters in depth, fitting of l_1-norm result to true model is superior, especially from $z = 1850$ m. In this figure, $\mathbf{m}$ stands for true model and $\mathbf{m}_0$ for initial model.

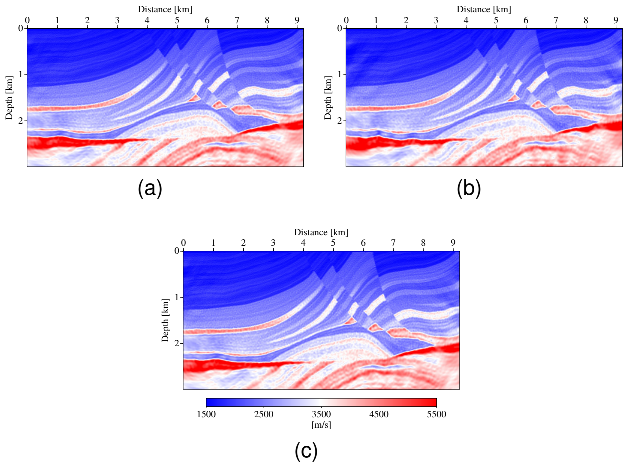

Figure 4: Estimated Marmousi model after inversion using noisy data as entry. Objective functionals are based on l_2-norm (a), l_1-norm (b) and cross-correlation of data vectors (c).

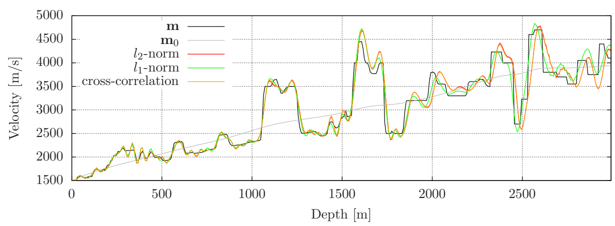

Figure 5: Velocity profile of estimated model using noisy data as entry for different objective functionals. Profile is located at $x = 6375$ m.

Uncertainties in source signature

In this experiment, observed data were modeled using $s^{\mathrm{t}}(t)$, a modified version of zero-phase Ricker wavelet, as source function. This is computed as $s^{\mathrm{t}}(t) = s^{\mathrm{i}}(t)e^{5.0t}$, where $s^{\mathrm{i}}(t)$ corresponds to the conventional form of Ricker wavelet and which is assumed as the estimated source signature in the iterative scheme (Figure 6). For visualization purposes, only 0.5 s of signal length is shown.

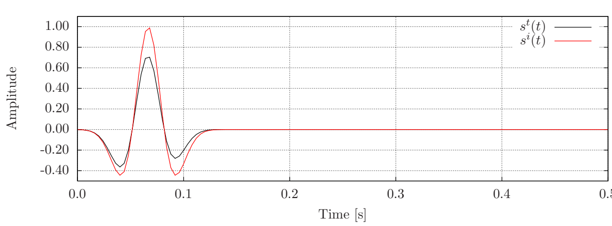

Figure 6: Source function used for modeling observed data $s^{\mathrm{t}}(t)$ and estimated signature used for inversion $s^{\mathrm{i}}(t)$.

Inversion result for objective functions based on l_1-norm and cross-correlation between data vectors are shown in Figures 7.b and 7.c respectively. In the case of l_2-norm, the scheme does not converge due to errors in the amplitude in source function used for inversion. We introduced the coefficient ρ, given by (Wang et al., 2012)

$$\rho = \frac{\langle \mathbf{d}_{\mathrm{cal}}, \mathbf{d}_{\mathrm{obs}} \rangle}{\|\mathbf{d}_{\mathrm{cal}}\|_2^2}, \tag{10}$$

in order to normalize the amplitude of calculated data vector before computing secondary virtual source. Therefore, the new form of data residual used for calculating adjoint-state variable wavefield is $\rho\mathbf{d}_{\mathrm{cal}} - \mathbf{d}_{\mathrm{obs}}$. ρ has the same form as the coefficient applied to synthetic data vector when computing secondary source for objective function defined in terms of cross-correlation (equation 8c). Solution for l_2-norm, obtained normalizing amplitudes in $\mathbf{d}_{\mathrm{cal}}$, is shown in Figure 7.a.

Estimated models (Figure 7) make clear the effects of errors in estimated source signature. Geometry of structures and velocity values are altered, especially at deeper parts. Nevertheless, all the three formulations were able to recover the most noticeable tendencies in model parameter values, being observable also in a depth profile (Figure 8). Considering that amplitude of secondary virtual source for l_1-norm is independent from the magnitude of the difference in amplitude of data vectors, this approach has a considerable stability faced with an erroneous approximation of source function. On the other hand, when cross-correlation based objective function is used, synthetic traces are scaled accordingly to its normalized inner product with observed traces, controlling the amplitude of secondary virtual source in the computation of adjoint-state wavefield (Dutta et al., 2014).

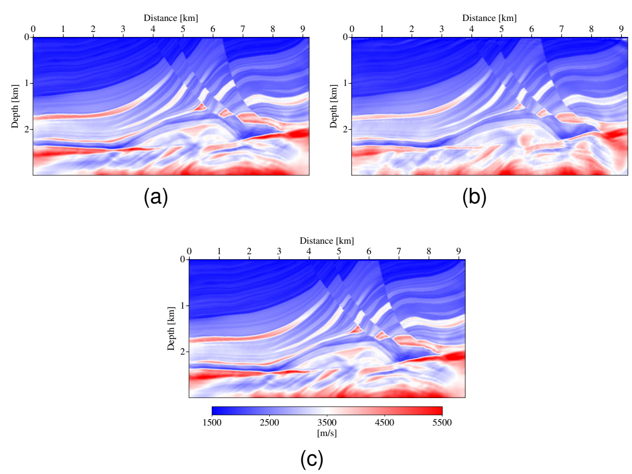

Figure 7: Estimated Marmousi model after inversion using an approximated source signature. Objective functionals are based on l_2-norm, normalizing secondary virtual source (a), l_1-norm (b) and cross-correlation of data vectors (c).

FWI, when guided by l_2-norm objective function, focuses on data amplitude matching, being sensitive to errors in source function assumed in iterative scheme, induced in this experiment. Solution estimated, applying the normalization of modeled data, is equivalent to that obtained by using cross-correlation objective function (Figure 8), in which scaling $\mathbf{d}_{\mathrm{cal}}$ is performed inherently.

The corresponding NRMS errors are 11.38%, 11.03% and 11.39%, for objective functions defined in terms of l_2-norm,

l_1-norm and cross-correlation, respectively. Both cases (Figures 5 and 8) show an outstanding agreement between the solutions obtained by using l_2-norm and correlation of data vectors in the formulation of objective function, reflecting that they are equivalent, at least in the scaled form of secondary virtual source.

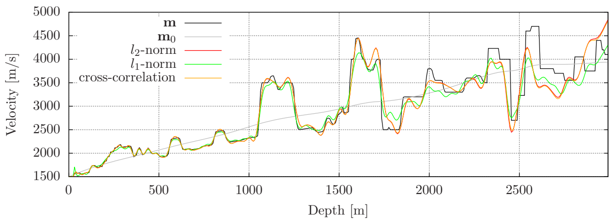

Figure 8: Velocity profile of estimated model using an approximated source signature for different objective functionals. Profile is located at $x = 6375$ m.

Conclusions

The characteristics and the performance of alternative formulations for FWI objective function, besides the conventionally used l_2-norm, are studied. Numerical synthetic examples make clear how the approaches based on l_1-norm and cross-correlation of data vectors are more stable when facing aleatory noisy affecting data and seismic source signature is not accurately estimated.

The behavior of objective functions based on l_2-norm and cross-correlation are equivalent, at least in the normalization that the last one performs on amplitude of secondary virtual source during gradient computation, making the scheme more robust when amplitudes of calculated data do not correspond with amplitude of observed data.

The successful application of the quasi-Newton L-BFGS algorithm shows that it can be applied to non-quadratic objective functions, without affecting the convergence of iterative scheme significantly. This conclusion had been stated for l_1-norm based objective functions and we extend to cross-correlation between data vectors.

Acknowledgements

This research was supported by CNPq, CAPES and INCT-GP/CNPq. The facility support from CPGG/UFBA is also acknowledged. Authors thank to Adriano W. G. dos Santos and Daniel E. Revelo for providing useful computer algorithms.

References

Brossier, R., S. Operto, and J. Virieux, 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: Geophysics, **75**, R37–R46.

Bube, K. P. and R. T. Langan, 1997, Hybrid l^1/l^2 minimization with applications to tomography: Geophysics, **62**, 1183–1195.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Dutta, G., M. Sinha, and G. T. Schuster, 2014, A cross-correlation objective function for least-squares migration and visco-acoustic imaging: SEG Technical Program Expanded Abstracts, 3985–3990.

Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: The Annals of Statistics, **1**, 799–821.

Klimm, B., 2013, Time domain full waveform inversion using ADI modeling: PhD thesis, Technischen Universität Kaiserslautern.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: theory and application, 206–220, SIAM.

Nocedal, J., 1980, Updating quasi-newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.

Nocedal, J. and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.

Pestana, R. C. and P. L. Stoffa, 2010, Time evolution of the wave equation using rapid expansion method: Geophysics, **75**, T121–T131.

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

———, 2005, Inverse problem theory and methods for model parameter estimation: Society for Industrial and Applied Mathematics.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Wang, C., D. Yingst, R. Bloor, and J. Leveille, 2012, Application of VTI waveform inversion with regularization and preconditioning to real 3D data: EAGE 74th Conference & Exhibition, W009.

Zhang, Y., L. Duan, and Y. Xie, 2015, A stable and practical implementation of least-squares reverse time migration: Geophysics, **80**, V23–V31.